

Large cliques in sparse random intersection graphs*

Mindaugas Bloznelis and Valentas Kurauskas

Vilnius University

2013-02-17

Abstract

Given positive integers n and m , and a probability measure P on $\{0, 1, \dots, m\}$ the random intersection graph $G(n, m, P)$ on vertex set $V = \{1, 2, \dots, n\}$ and with attribute set $W = \{w_1, w_2, \dots, w_m\}$ is defined as follows. Let S_1, S_2, \dots, S_n be independent random subsets of W such that for any $v \in V$ and any $S \subseteq W$ we have $\mathbb{P}(S_v = S) = P(|S|)/\binom{m}{|S|}$. The edge set of $G(n, m, P)$ consists of those pairs $\{u, v\} \subseteq V$ for which $S_u \cap S_v \neq \emptyset$.

We study the asymptotic order of the clique number $\omega(G(n, m, P))$ in random intersection graphs with bounded expected degrees. For instance, in the case $m = \Theta(n)$ we show that if the vertex degree distribution is power-law with exponent $\alpha \in (1; 2)$, then the maximum clique is of a polynomial size, while if the variance of the degrees is bounded, then the maximum clique has $\frac{\ln n}{\ln \ln n}(1 + o(1))$ vertices whp.

In each case we give a polynomial algorithm which finds a clique of size $\omega(G(n, m, P))(1 - o(1))$ whp. One of ingredients of our proofs is a result of Alon et al [1] in Ramsey theory.

keywords: clique, random intersection graph, greedy algorithm, sparse network, power-law

corresponding author:

Valentas Kurauskas, Faculty of Mathematics and Informatics, Didlaukio 47,
LT-08303 Vilnius, Lithuania

email:

Mindaugas Bloznelis: mindaugas.bloznelis@mif.vu.lt

Valentas Kurauskas: valentas@gmail.com

*Part of this work was supported by the Research Council of Lithuania (MIP-053/2011).

1 Introduction

Bianconi and Marsili observed in 2006 that empirical ‘scale-free’ networks can have very large cliques; they gave an argument suggesting that the rate of divergence is polynomial if the degree variance is unbounded [2]. In a more precise analysis, Janson, Łuczak and Norros [9] gave exact asymptotics for the clique number in a power-law random graph model where edge probabilities are proportional to the product of weights of their endpoints. However, with conditionally independent edges, the random graphs of [9] do not have the clustering property which may be important in explaining the presence of large cliques in real-world networks.

The random intersection graph model was introduced by Karoński, Scheinerman and Singer-Cohen in 1999 [11] and further generalised by Godehardt and Jaworski [8] and others. With appropriate parameters, it yields graphs that are sparse [5], have a positive clustering coefficient [6, 4] and assortativity [4]. In the introductory paper [11], thresholds for small (constant-sized) cliques were determined; the asymptotic distribution of the number of small cliques was studied in [14]. We provide asymptotics for the size of the largest clique in random intersection graphs with linear number of edges.

For a sequence $(G(n), n = 1, 2, \dots)$ of random intersection graphs with $G(n) = G(n, m, P)$ and $m = m(n)$, $P = P(n)$ we let $X(n)$ denote a random variable distributed according to $P(n)$ and define $Y(n) := \sqrt{\frac{n}{m}} X(n)$. If not stated otherwise the limits below will be taken as $n \rightarrow \infty$. In this paper we use the standard notation $o()$, $O()$, $\Omega()$, $\Theta()$, $o_P()$, $O_P()$, see, for example, [10]. We will consider sequences $(G(n))$ for which

$$\mathbb{E} Y(n) = O(1). \quad (1)$$

This condition ensures that the resulting graphs are sparse, that is, the expected number of edges in $G(n)$ is $O(n)$.

A function $L : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is called *slowly varying* if $\lim_{x \rightarrow \infty} L(tx)/L(x) = 1$ for any $t > 0$. For positive sequences (a_n) , (b_n) we write $a_n \sim b_n$ if $a_n/b_n \rightarrow 1$ as $n \rightarrow \infty$. We say that $(G(n))$ is *power-law with index α* if there are positive numbers α , ϵ_0 and a slowly varying function L such that for each sequence x_n with $n^{1/2-\epsilon_0} \leq x_n \leq n^{1/2+\epsilon_0}$ we have

$$\mathbb{P}(Y(n) \geq x_n) \sim L(x_n) x_n^{-\alpha}. \quad (2)$$

Our first theorem shows that when $\alpha \in (1; 2)$ the order of the maximum clique in $G(n, m, P)$ is as in the considerably simpler model of [9].

Theorem 1.1 *Let $(G(n), n = 1, 2, \dots)$, $G(n) = G(n, m(n), P(n))$ be a sequence of random power-law intersection graphs with index $\alpha \in (1; 2)$. Suppose that $\mathbb{E} Y(n) = O(1)$ and there is $\beta > \max(2 - \alpha, \alpha - 1)$ such that $m = m(n) = \Omega(n^\beta)$. Let L be a slowly varying function given in (2). Then the clique number of $G(n)$ is*

$$\omega(G(n)) = K(n)(1 + o_P(1))$$

where

$$K(n) = (1 - \alpha/2)^{-\alpha/2} L\left((n \ln n)^{1/2}\right) n^{1-\alpha/2} (\ln n)^{-\alpha/2}.$$

The set of edges of a random intersection graph $G = G(n, m, P)$ with attribute set W is a union of *monochromatic* cliques $T(w) = \{v \in [n] : w \in S_v\}$. We denote the size of a largest monochromatic clique by $\omega'(G)$. Clearly, the clique number of G is at least $\omega'(G)$.

Our second theorem shows that the largest clique in a sparse random intersection graph with variance of Y bounded is a monochromatic clique (plus possibly a few extra vertices). Denote $x \vee y = \max(x, y)$.

Theorem 1.2 *Let $(G(n), n = 1, 2, \dots)$, $G(n) = G(n, m(n), P(n))$ be a sequence of random intersection graphs. Suppose that $\mathbb{E}Y(n) = O(1)$, $m(n) \rightarrow \infty$ and $\text{Var}(Y(n)) = O(1)$. Then*

$$\mathbb{E} (\omega(G(n)) - \omega'(G(n)))^2 = O(1).$$

If, in addition, there is a positive sequence (ϵ_n) such that $\epsilon_n \rightarrow 0$ and

$$n\mathbb{P}(Y(n) > \epsilon_n n^{1/2}) \rightarrow 0 \tag{3}$$

then

$$\mathbb{P}(\omega(G(n)) \leq C \vee (\omega'(G(n)) + 3)) \rightarrow 1$$

for an absolute constant C .

Let us recall the maximum load problem. Consider the balls into bins model where each of N balls is thrown into one of m bins uniformly and independently at random. Let $M(N, m)$ be the maximum number of balls contained in any of the bins. The asymptotics of $M(N, m)$ are well known, see, e.g., Section 6 of Kolchin et al [12].

Denote by $d_{TV}(X, Y)$ the total variation distance between random variables X and Y (that is, $d_{TV}(X, Y) = \sup |\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)|$ over all Borel sets A). Our next result supplements Theorem 1.2 by providing asymptotics for the size of a largest monochromatic clique.

Theorem 1.3 *Let $(G(n), n = 1, 2, \dots)$, $G(n) = G(n, m(n), P(n))$ be a sequence of random intersection graphs. Let $Y = Y(n)$ be defined as in (1). Suppose¹ that $\mathbb{E}Y = \Theta(1)$, $\text{Var}(Y) = O(1)$ and $m(n) \rightarrow \infty$. Then*

$$d_{TV}(\omega'(G(n)), M(\lfloor \mathbb{E}Y(mn)^{1/2} \rfloor, m)) \rightarrow 0.$$

For example, if $m(n) = \Theta(n)$, Theorem 1.2 and Theorem 1.3 together with the asymptotics for M (Theorem II.6.1 of [12]), imply that

$$\omega(G(n)) = \frac{\ln n}{\ln \ln n} (1 + o_P(1)).$$

¹A small technical remark: Theorem 1.3 does not make the assumption (3), but asks that $\mathbb{E}Y$ is bounded away from zero.

In contrast, for a constant $c > 0$, the Erdős-Rényi random graph has $\omega(G(n, c/n)) \leq 3$ whp and in the model of [9] with corresponding parameters the maximum clique has at most 4 vertices.

The condition (3) is not very restrictive. It can be replaced by a stronger requirement that both a) Y converges in distribution to a random variable Y^* with $\text{Var}(Y^*) < \infty$, and b) $\text{Var}(Y)$ converges to $\text{Var}(Y^*)$, see, e.g., [5].

Each of our main results (Theorem 1.1 and Theorem 1.2) have corresponding simple polynomial algorithms that construct a clique of the optimal order whp. For power-law random intersection graphs with $\alpha \in (1; 2)$, we may use a greedy algorithm considered in [9]. Sort the vertices of a graph in descending order according to their degree. Traverse vertices in that order and ‘grow’ a clique, by adding a vertex if it is connected to each vertex in the current clique.

For random intersection graphs corresponding to Theorem 1.2, we propose another simple algorithm based on the fact that the largest clique whp contains two vertices which share only one attribute. For each pair of connected vertices, take any maximal clique formed by that pair and their common neighbours. Output the biggest maximal clique found in this way. More details and analysis of each of the algorithms are given in Section 4 below.

An empirical graph may be assumed to be distributed as a random intersection graph, but information about the subset size distribution may not be available. Suppose $(G(n), n = 1, 2, \dots)$ is a sequence of random intersection graphs. Instead of the condition (2) for the tails of the normalised subset sizes, $Y(n)$, we may consider a similar condition for the degree $D_1(n) = d_{G(n)}(1)$ of vertex 1 in $G(n)$: there are constants $\alpha' > 1, \epsilon' > 0$ and a slowly varying function $L'(x)$ such that for any sequence t_n with $n^{1/2-\epsilon'} \leq t_n \leq n^{1/2+\epsilon'}$

$$\mathbb{P}(D_1(n) \geq t_n) \sim L'(t_n)t_n^{-\alpha'}. \quad (4)$$

The following lemma shows that, subject to an additional assumption, there is equivalence between these two conditions.

Lemma 1.4 *Let $(G(n), n = 1, 2, \dots)$, $G(n) = G(n, m(n), P(n))$ be a sequence of random intersection graphs, let $D_1 = D_1(n)$ be the degree of vertex 1 in $G(n)$ and let $Y = Y(n)$ be defined as in (1). Assume there is $\epsilon > 0$ such that*

$$\mathbb{E} Y \mathbf{1}_{Y \geq n^{1/2-\epsilon}} \rightarrow 0. \quad (5)$$

Suppose either $(\mathbb{E} Y)^2$ or $\mathbb{E} D_1$ converges to a positive number.

Then both limits exist and are equal. Furthermore, the condition (4) holds if and only if (2) holds. In that case, $\alpha' = \alpha$ and $L'(t) \sim d^{\alpha/2} L(t)$, where $d = \lim \mathbb{E} D_1 = \lim (\mathbb{E} Y)^2$.

Thus, if we make the assumption (5) then we can use only the information about the degree distribution in Theorem 1.1.

Similarly, in Theorem 1.2 the condition that $\mathbb{E} Y^2$ is bounded may be replaced by a condition that the degree variance is bounded.

Lemma 1.5 *Let $(G(n), n = 1, 2, \dots)$, $G(n) = G(n, m(n), P(n))$ be a sequence of random intersection graphs, let $D_1 = D_1(n)$ be the degree of vertex 1 in $G(n)$ and let $Y = Y(n)$ be defined as in (1). Assume there is a sequence $\epsilon_n \rightarrow 0$ such that*

$$\mathbb{E} Y^2 \mathbf{I}_{Y > \epsilon_n n^{1/2}} \rightarrow 0 \quad (6)$$

and either $\mathbb{E} Y = \Theta(1)$ or $\mathbb{E} D_1 = \Theta(1)$. Then

$$\mathbb{E} D_1 = (\mathbb{E} Y)^2 + o(1) \quad (7)$$

$$\text{Var}(D_1) = \mathbb{E} Y (\text{Var}(Y) + 1) + o(1). \quad (8)$$

The plan of the rest of this paper is as follows. In Section 2 we study sparse random power-law intersection graphs with index $\alpha \in (1; 2)$, introduce the result on “rainbow” cliques in extremal combinatorics (Lemma 2.8) and prove Theorem 1.1. Next, in Section 3 we relate our model to the balls and bins model and prove Theorem 1.2. In Section 4 we present and analyse algorithms for finding large cliques in $G(n, m, P)$. Then, in Section 5 we prove Lemmas 1.4 and 1.5. Finally, in the last section we give some concluding remarks.

2 Random power-law intersection graphs

2.1 Preliminaries

Let $\{G(n), n = 1, 2, \dots\}$ be a sequence of power-law random intersection graphs with index $\alpha \in (1; 2)$, where $G(n) = G(n, m, P)$, $m = m(n)$ and $P = P(n)$. Assume that (1) holds. Let $L : R_+ \rightarrow R_+$ be the slowly varying function given in (2).

Let ϵ_1 be a small positive constant and define

$$\theta_1 = \theta_1(n) = m^{1/2} n^{-\epsilon_1}; \quad \theta_2 = \theta_2(n) = ((1 - \alpha/2)m \ln n + m e_1)^{1/2},$$

where² $e_1 = e_1(n) = \max(0, \ln L((n \ln n)^{1/2}))$. For $v \in V(G(n))$ let $X_v(n) = |S_v(n)|$ be the size of random subset corresponding to v . For each n define random sets

$$\begin{aligned} V_0 &= V_0(n) = \{v \in V(G(n)) : X_v < \theta_1\}; \\ V_1 &= V_1(n) = \{v \in V(G(n)) : \theta_1 \leq X_v \leq \theta_2\}; \\ V_2 &= V_2(n) = \{v \in V(G(n)) : \theta_2 < X_v\}. \end{aligned}$$

Also write $G_i = G_i(n) = G(n)[V_i]$ for $i = 1, 2, 3$. The main result follows from three lemmas about each of the parts V_i . Let $K = K(n)$ be as in Theorem 1.1. The first lemma gives a lower bound for the clique number of $G(n)$.

Lemma 2.1 *For any $m = m(n)$*

$$\omega(G_2) = |V_2|(1 - o_P(1)) = K(1 - o_P(1)).$$

²To understand the idea of the proof one can assume $L(x) = 1$ and $e_1 = 0$.

The next two lemmas provide the upper bound.

Lemma 2.2 *Suppose there is $\beta > \alpha - 1$ such that $m = \Omega(n^\beta)$. If $\epsilon_1 < \frac{\beta}{6}$ then there is $\delta > 0$ such that*

$$\mathbb{P}\left(\omega(G_0) \geq n^{1-\alpha/2-\delta}\right) \rightarrow 0.$$

Lemma 2.3 *Suppose there is $\beta > 2 - \alpha$ such that $m = \Omega(n^\beta)$. If $\epsilon_1 < \frac{\beta-2+\alpha}{24}$ then*

$$\omega(G_1) = o_P(K).$$

Proof of Theorem 1.1 We can take $\epsilon_1 > 0$ so that $\epsilon_1 < \min(\frac{\alpha-1}{6}, \frac{\beta-2+\alpha}{24})$. By Lemma 2.1, $\omega(G) \geq \omega(G_2) = K(1 - o_P(1))$. Since $\omega(G) \leq \omega(G_0) + \omega(G_1) + \omega(G_2)$ we have $\omega(G) \leq K(1 + o_P(1))$ by Lemmas 2.2 and 2.3. \square

We will often use a random intersection graph $G(n, m, P)$ with P the binomial distribution with parameters m and p . This model is called the *random binomial intersection graph* and denoted $G(n, m, p)$.

In $G(n, m, p)$ the events $w \in S_v$ are independent for all pairs (w, v) , $w \in W$, $v \in V = [n]$, and occur with probability p . For a random variable B with distribution $\text{Binom}(m, p)$ and any positive $\epsilon < \frac{3}{2}$, we have by a Chernoff bound (see [13]):

$$\mathbb{P}(|B - mp| > \epsilon mp) \leq 2e^{-\frac{1}{3}\epsilon^2 mp}. \quad (9)$$

Therefore if $(G(n, m, p), n = 1, 2, \dots)$ is a sequence of random binomial intersection graphs with $m = m(n)$, $p = p(n)$, then the sizes $X_v = X_v(n)$ of the random subsets of $G(n, m, p)$ satisfy

$$\mathbb{P}(\exists v \in [n] : |X_v - mp| > y) \leq n\mathbb{P}(|X_v - mp| > y) \rightarrow 0 \quad (10)$$

for any $y = y(n)$ such that $y/\sqrt{mp \ln n} \rightarrow \infty$.

2.2 Large sets (the graph G_2)

In this section we use ideas from [9] to give a lower bound on the clique size for a sequence of sparse random intersection graphs.

We first note the following technical facts.

Lemma 2.4 *Suppose $a = a_n, b = b_n$ are sequences of positive reals such that $0 < \ln 2b + 2a \rightarrow +\infty$. Let z_n be the positive root of*

$$a - \ln z - bz^2 = 0. \quad (11)$$

Then $z_n \sim \sqrt{\frac{2a + \ln(2b)}{2b}}$.

Proof Changing the variables $t = 2bz^2$ we get

$$t + \ln(t) = 2a + \ln(2b).$$

From the assumption it follows that $t + \ln t \sim t$ and therefore $z_n \sim \sqrt{\frac{2a + \ln(2b)}{2b}}$. \square

Lemma 2.5 ([7]) *For any slowly varying function L we have $\ln L(x) = o(\ln x)$ as $x \rightarrow \infty$.*

Proof of Lemma 2.1 Write $N = |V_2|$ and let

$$v^{(1)}, v^{(2)}, \dots, v^{(N)}$$

be the vertices of V_2 listed in a descending order of their set sizes, i.e. $|S_{v^{(i)}}| \leq |S_{v^{(i-1)}}|$ for $i = 2, \dots, N$.

Consider a greedy algorithm for finding a clique in G proposed by Janson, Łuczak and Norros [9] (see also Section 4). Let $A^0 = \emptyset$. In the step $i = 1, 2, \dots, N$ let $A^i = A^{i-1} \cup \{v^{(i)}\}$ if $v^{(i)}$ is incident to each of the vertices $v^{(j)}$, $j = 1, \dots, i-1$. Otherwise let $A^i = A^{i-1}$. This algorithm produces a clique H on the set of vertices A^N , and H demonstrates that $\omega(G_2) \geq |A^N|$.

Let L be the set of vertices that failed to be added to A^N , i.e. $L = V_2 \setminus A^N$. We will show that

$$|L|/N = o_P(1)$$

and

$$N = (1 - \alpha/2)^{-\alpha/2} L \left((n \ln n)^{1/2} \right) (\ln n)^{-\alpha/2} n^{1-\alpha/2} (1 - o_P(1)).$$

Since by Markov's inequality $\mathbb{P}(|L|/N > \epsilon) \leq \frac{\mathbb{E}(|L|/N)}{\epsilon}$ it is enough to show that

$$\mathbb{E}(|L|/N) \rightarrow 0.$$

Write $\theta = \theta_2$. Let p_1 be the probability that two random θ -sets from $W = [m]$ do not intersect. Conditionally on N , the number of vertices in L is at most the number of pairs in $x, y \in V_2$ where S_x and S_y do not intersect.

Therefore by the first moment method

$$\mathbb{E}(|L|/N) = \mathbb{E} \mathbb{E}(|L|/N | N) \leq \mathbb{E} \left(\frac{\binom{N}{2} p_1}{N} \middle| N \right) \leq \frac{\mathbb{E} N p_1}{2}.$$

Now

$$p_1 = \frac{\binom{m-\theta}{\theta}}{\binom{m}{\theta}} \leq \left(1 - \frac{\theta}{m} \right)^\theta \leq e^{-\theta^2/m}$$

and $N \sim \text{Binom}(n, q)$ where $q = \mathbb{P}(X_n > \theta)$. Using (2) we have

$$\begin{aligned} \mathbb{E} N &= nq = n\mathbb{P} \left((m/n)^{1/2} Y_n > \theta \right) \\ &\sim L \left((n/m)^{1/2} \theta \right) n^{1-\alpha/2} m^{\alpha/2} \theta^{-\alpha} \\ &\sim (1 - \alpha/2)^{-\alpha/2} L(\sqrt{n \ln n}) (\ln n)^{-\alpha/2} n^{1-\alpha/2}. \end{aligned}$$

Here we used equivalences $L((n/m)^{1/2} \theta) \sim L(\sqrt{n \ln n})$ and $\ln L(\sqrt{n \ln n}) = o(\ln n)$ which follow from Lemma 2.5

By (10) we have $N = EN(1 - o_P(1))$. It is straightforward to check that for some constant c we have $\mathbb{E} N p_1 \leq c(\ln n)^{-\alpha/2} \rightarrow 0$. This completes the proof.

Let us briefly explain the intuition for the choice of θ . For simplicity assume $L(x) = 1$ so that $e_1 = 0$. Could the same method yield a bigger clique if θ_2 is smaller? It turns out that the given value for θ_2 is (asymptotically) optimal. It can be obtained by solving

$$n^{1-\alpha/2} m^{\alpha/2} \theta^{-\alpha} e^{-\theta^2/m} = 1$$

or

$$\alpha^{-1} \ln n + 1/2 \ln(m/n) - \ln \theta - \frac{\theta^2}{\alpha m} = 0 \quad (12)$$

which is of the form (11) where $a = \alpha^{-1} \ln n + 1/2 \ln(m/n)$ and $b = (\alpha m)^{-1}$.

We have that $be^{2a} = \alpha^{-1} n^{\frac{2}{\alpha}-1} \rightarrow +\infty$ and by Lemma 2.4 the solution of (12) satisfies

$$\theta \sim \sqrt{\frac{(2/\alpha) \ln n - \ln(n/m) + \ln(2/\alpha m)}{2/\alpha m}} \sim \sqrt{(1 - \alpha/2)m \ln n}.$$

□

2.3 Small sets (the graph G_0)

In this section we consider the graph G_0 . This graph has $n(1 - o_P(1))$ vertices and all the sets S_v are of size at most $m^{1/2} n^{-\epsilon_1}$. We say that an intersection graph $G = G(V, W)$ contains a rainbow K_h if there is a subgraph $H \subseteq G$ on h vertices and an injective map, assigning each edge $xy \in E(H)$ an attribute $w_{xy} \in S_x \cap S_y \subseteq W$. The next lemma follows directly from the results of [11].

Lemma 2.6 *Let $(G(n), n = 1, 2, \dots)$, $G(n) = G(n, m, p)$, $m = m(n)$, $p = p(n)$ be a sequence of binomial random intersection graphs, let h be a positive integer and suppose that $pn^{1/(h-1)}m^{1/2} \rightarrow a \in \{0, 1\}$. Then*

$$\mathbb{P}(G \text{ contains a rainbow } K_h) \rightarrow a.$$

Proof The case $a = 1$ follows from Claim 2 of [11]. For the case $a = 0$ we have by the first moment method

$$\begin{aligned} \mathbb{P}(G \text{ contains a rainbow } K_h) &\leq \binom{n}{h} (m)_{(h)} p^{2\binom{h}{2}} \\ &\leq \left(n^{1/(h-1)} m^{1/2} p \right)^{h(h-1)} \rightarrow 0 \end{aligned}$$

when $n^{1/(h-1)} m^{1/2} p \rightarrow 0$. □

In the next lemma we give a crude upper bound for $\omega'(G)$.

Lemma 2.7 *Let $(G(n), n = 1, 2, \dots)$, $G(n) = G(n, m, P)$, $m = m(n)$, $P = P(n)$ be a sequence of random intersection graphs such that $\mathbb{E} Y(n) = O(1)$. Suppose $(G(n))$ is power-law with index $\alpha \in (1; 2)$ and there is $\beta > \alpha - 1$ such that $m = \Omega(n^\beta)$. Then there is a constant $\delta > 0$ such that $\omega'(G(n)) \leq n^{1-\alpha/2-\delta}$ whp.*

Proof Write $V = V(G(n)) = [n]$, $W = W(G(n)) = \{w_1, \dots, w_m\}$ and let $X = X(n)$ and $Y = Y(n)$ be defined as in (1). For a fixed key $w \in W$ and a vertex $v \in V$

$$\mathbb{P}(w \in S_v) = \sum_{k=0}^{\infty} \frac{k}{m} \mathbb{P}(|S_v| = k) = \frac{\mathbb{E} X}{m} = \frac{\mathbb{E} Y}{\sqrt{mn}}.$$

Since

$$|T_w| \sim \text{Binom}\left(n, \frac{\mathbb{E} Y}{\sqrt{mn}}\right) \quad (13)$$

we have, for any positive integer k

$$\mathbb{P}(|T_w| \geq k) \leq \binom{n}{k} \left(\frac{\mathbb{E} Y}{\sqrt{mn}}\right)^k \leq \left(\frac{en\mathbb{E} Y}{k\sqrt{mn}}\right)^k \leq \left(\frac{c_1}{k} \sqrt{\frac{n}{m}}\right)^k$$

for $c_1 = e \sup_n \mathbb{E} Y$. Therefore

$$\mathbb{P}(\omega'(G(n)) \geq k) \leq m \left(\frac{c_1}{k} \sqrt{\frac{n}{m}}\right)^k.$$

Fix δ with $0 < \delta < \min((\alpha - 1 - \beta)/4, 1 - \alpha/2, \beta/2)$. We have

$$\begin{aligned} \mathbb{P}(\omega'(G(n)) \geq n^{1-\alpha/2-\delta}) &\leq m \left(c_1 n^{\alpha/2-1/2+\delta} m^{-1/2}\right)^{\lceil n^{1-\alpha/2-\delta} \rceil} \\ &= m^{1-(\delta/\beta)\lceil n^{1-\alpha/2-\delta} \rceil} \left(c_1 n^{\alpha/2-1/2+\delta} m^{-1/2+\delta/\beta}\right)^{\lceil n^{1-\alpha/2-\delta} \rceil} \rightarrow 0 \end{aligned}$$

since $m \rightarrow \infty$, $n^{1-\alpha/2-\delta} \rightarrow \infty$ and for large enough n

$$n^{\alpha/2-1/2+\delta} m^{-1/2+\delta/\beta} \leq n^{\alpha/2-1/2+\delta-\beta/2+\delta} = n^{(\alpha-1-\beta+4\delta)/2} \rightarrow 0.$$

□

The last and the most important fact we need relates the maximum clique size with the maximum rainbow clique size in an intersection graph. An edge-colouring of a graph is called t -good if each colour appears at most t times at each vertex. Given an edge-coloured graph G we say that G contains a rainbow copy of H if G contains a subgraph \tilde{H} isomorphic to H such that no two edges of \tilde{H} have the same colour.

Lemma 2.8 ([1]) *There is a constant c such that every t -good coloured complete graph on more than $\frac{ct h^3}{\ln h}$ vertices contains a rainbow copy of K_h .*

Proof of Lemma 2.2 Define $p = p(n) = m^{-1/2}n^{-\epsilon_1} + m^{-2/3}$ and fix an integer $h > 1 + \frac{1}{\epsilon_1}$. Then

$$\mathbb{P}(G_0 \text{ contains a rainbow } K_h) \rightarrow 0. \quad (14)$$

Indeed, since $\epsilon_1 < \frac{\beta}{6}$, by (10) each random subset in $G(n, m, p)$ has size at least θ_1 whp and the probability that G_0 contains a rainbow K_h can be bounded by

$$\mathbb{P}(G(n, m, p) \text{ contains a rainbow } K_h) + o(1).$$

Hence (14) follows by Lemma 2.6 since $n^{1/(h-1)}m^{1/2}p \rightarrow 0$.

We have

$$\mathbb{P}(\omega(G_0) \geq k) \leq \mathbb{P}(\omega'(G(n)) > t) + \mathbb{P}(\omega(G_0) \geq k \mid \omega'(G(n)) \leq t).$$

By Lemma 2.7 there is $\delta > 0$ such that for $t = n^{1-\alpha/2-\delta}$

$$\mathbb{P}(\omega'(G(n)) > t) \rightarrow 0.$$

By (14), the probability that G_0 contains a rainbow clique of size h conditional on $\omega'(G(n)) \leq t$ tends to 0. Therefore by Lemma 2.8 the probability that G_0 contains a rainbow clique of size at least $\frac{ct h^3}{\ln h} = O(n^{1-\alpha/2-\delta})$ also tends to 0. [To see that we can apply Lemma 2.8 to an intersection graph G , colour each edge $xy \in E(G)$ by an arbitrary element of $S_x \cap S_y$. If $\omega'(G) \leq t$ then clearly each colour appears at each vertex at most t times.] \square

2.4 Intermediate sets (the graph G_1)

We start with a version of the result of Erdős and Rényi about the maximum clique (see, for example, [10]). By the *number of cliques* of size t in a graph G we mean the number of subsets $S \subseteq V(G)$ of size t which induce a clique in G .

Lemma 2.9 *Consider a sequence of Erdős and Rényi random graphs $(G(n, p))$ $n = 1, 2, \dots$ with $p = p(n) \rightarrow 1$. Let $K = K(n) = \frac{2 \ln n}{1-p}$ and suppose $r = r(n)$ satisfies $r = o(K^2)$.*

Then there are positive $\delta = \delta(n)$ and $\epsilon = \epsilon(n)$ such that $\delta, \epsilon \rightarrow 0$ and for any n and any sequence of graphs (R_n) with $V(R_n) = [n]$ and $|E(R_n)| \leq r(n)$ the number $X(n)$ of cliques of size $\lfloor K(1 + \delta(n)) \rfloor$ in $G(n, p) \cup R_n$ satisfies

$$\mathbb{E} X(n) \leq \epsilon(n).$$

Proof Write $h = 1 - p$ and pick a positive sequence δ so that $\delta \rightarrow 0$ and $\ln^{-1} n + h + \frac{r}{K^2} = o(\delta)$. Let $a = \lceil K(1 + \delta) \rceil$. We have

$$\mathbb{E} X(n) \leq \binom{n}{a} p^{\binom{a}{2}-r} \leq \left(\frac{en}{a}\right)^a p^{\frac{a(a-1)}{2}-r}.$$

But

$$\begin{aligned}
\ln(en/a) + \left(\frac{a-1}{2} - \frac{r}{a}\right) \ln p &= \ln(en/a) + \left(\frac{a-1}{2} - \frac{r}{a}\right) (-h + O(h^2)) \\
&\leq \ln n - \frac{ah}{2} + O\left(ah\left(\frac{r}{a^2} + h\right)\right) \\
&\leq \ln n \left(-\delta + O\left(\frac{r}{a^2} + h\right)\right) \rightarrow -\infty,
\end{aligned}$$

which completes the proof. \square

We will need the following lemma.

Lemma 2.10 *Let $\mathcal{F} = \{A_1, \dots, A_k\}$ be a family of sets from $\{1, \dots, n\}$. Let S be a random subset of $\{1, \dots, n\}$ of size d . If $|A_1| = a_1, \dots, |A_k| = a_k$ and $\sum_{i=1}^k a_i \leq n$ then*

$$\mathbb{P}(\{S \cap A_1, \dots, S \cap A_k\} \text{ has a system of distinct representatives}) \quad (15)$$

is maximised (subject to $|A_i| = a_i, i = 1, \dots, k$) when $\{A_i\}$ are independent.

Proof Call any of $\binom{n}{d}$ possible outcomes c for S a configuration and let \mathcal{C} be the family of all configurations. Given a configuration $c \in \mathcal{C}$, colour an element $x \in [n]$ black if $x \in c$ and colour it white otherwise. Given $\mathcal{F} = \{A_1, \dots, A_k\}$ let $\mathcal{C}_{DR}(\mathcal{F})$ be the set of all configurations c where $c \cap \mathcal{F} = \{c \cap A_1, \dots, c \cap A_k\}$ has a system of distinct representatives.

Suppose \mathcal{F} is a family of sets that maximises (15), and suppose without loss of generality that there is an element $x \in [n]$ and an index $i \in \{2, \dots, k\}$ such that $x \in A_1 \cap A_i$. Since $\sum_{i=1}^k |A_i| \leq n$ there is an element y in the complement of $\bigcup_{A \in \mathcal{F}} A$.

Now let $A'_1 = A_1 \setminus \{x\} \cup \{y\}$ and consider another family of sets $\mathcal{F}' = \{A'_1, A_2, \dots, A_k\}$.

Observe that the family of configurations $\mathcal{C} = \mathcal{C}_{DR}(\mathcal{F}) \setminus \mathcal{C}_{DR}(\mathcal{F}')$ has the following property: for each $c \in \mathcal{C}$ the element x is black, the element y is white and it is not possible to find a set of distinct representatives for $c \cap \mathcal{F}$ so that A is matched with an element other than x (otherwise excluding x from A and including y would not matter). Now, given a configuration c , let c_{xy} be a configuration with the colours of x and y swapped. Suppose $c \in \mathcal{C}$. Then $c_{xy} \notin \mathcal{C}_{DR}(\mathcal{F})$ and there is a set of distinct representatives for sets $c \cap (\mathcal{F} \setminus A_1)$ which does not use x . So $c_{xy} \in \mathcal{C}_{DR}(\mathcal{F})$, because y is black in c_{xy} and can be matched with A_1 . Thus each ‘bad’ configuration $c \in \mathcal{C}$ is compensated by a unique ‘good’ configuration $c_{xy} \in \mathcal{C}_{DR}(\mathcal{F}) \setminus \mathcal{C}_{DR}(\mathcal{F}')$. This shows that $|\mathcal{C}_{DR}(\mathcal{F}')| \geq |\mathcal{C}_{DR}(\mathcal{F})|$ and completes the proof. \square

Let $G = G(V, W)$ be an intersection graph, let t be a positive integer and let R be any subset of edges of the complete graph K_n . We denote by $\text{Rainbow}(G, R, t)$ the event that in the graph $G + R$ there is a clique H of size t where each edge $xy \in E(H) \setminus R$ can be represented by an attribute $w_{xy} \in S_x \cap S_y$ so that the resulting map from $E(H) \setminus R$ to W is injective. For two functions $f, g : \mathbb{N} \rightarrow \mathbb{R}_+$ we write $f \ll g$ if $f(n)/g(n) \rightarrow 0$.

Lemma 2.11 *Let $(G(n, m, p), n = 1, 2, \dots)$ be a sequence of random binomial intersection graphs where $m = m(n)$, $p = p(n)$. Suppose $K = K(n)$ and $r = r(n)$ are positive and $K = 2e^{mp^2} \ln n \leq n/2$, $mp^2 \rightarrow \infty$, $\frac{\ln n}{mp} \rightarrow 0$, $Kp \rightarrow 0$ and $r = o(K^2)$.*

Then there are positive sequences $(\epsilon_n), (\delta_n)$ such that $\epsilon_n, \delta_n \rightarrow 0$ and for any $(R(n), n = 1, 2, \dots)$ with $R = R(n) \subseteq E(K_n)$ and $|R(n)| \leq r(n)$

$$\mathbb{P}(\text{Rainbow}(G(n, m, p), R, K(1 + \delta_n))) \leq \epsilon_n, \quad n = 1, 2, \dots$$

Proof Let $x = x(n)$ be positive and such that

$$px \rightarrow 0, \quad \frac{x}{mp} \rightarrow 0 \text{ and } \sqrt{mp \ln n} \ll x.$$

Observe that such an x always exists since using the conditions of the lemma

$$\sqrt{mp \ln n} \ll mp \quad \text{and} \quad \sqrt{mp \ln n} = p^{-1} \sqrt{mp^3 \ln n} \ll p^{-1} \sqrt{pK} \ll p^{-1}.$$

Now consider n as fixed and let $M = mp + x$. For the graph $G = G(n, m, p)$ define random sets \bar{S}_v , $v \in V$ by

$$\bar{S}(v) = \begin{cases} S_v, & \text{if } |S_v| \leq M, \\ \text{the first } M \text{ elements of } S_v, & \text{otherwise.} \end{cases}$$

Fix $t \in [K; 2K]$. Let $T = \{v_1, \dots, v_t\}$ be any fixed subset of $V(G)$, let R_T be the edges in R with both ends in T .

For $i \in \{1, \dots, t\}$ we consider the event that the family of sets $\{\bar{S}_{v_j} \cap S_{v_i} \mid j = 1, \dots, i-1, v_j v_i \notin R\}$ has a set of distinct representatives. Denote this event by $A_T(i)$ and define

$$A_T := \bigcap_{i=1}^t A_T(i).$$

We will prove below that whenever n is large enough

$$\mathbb{P}(A_T) \leq (1 - (1 - p)^M)^{\binom{t}{2} - |R_T|}. \quad (16)$$

If $p' = 1 - (1 - p)^M$ and $G' = G(n, p')$ is the random Erdős-Rényi graph then the quantity on the right is the probability that the subgraph induced on T is complete in $G' + R$. Let X_t be the number of t -cliques in $G' + R$.

By Lemma 2.9 there are positive sequences $(\epsilon_n), (\delta_n)$ such that $\epsilon_n, \delta_n \rightarrow 0$ and for $K' = \frac{2 \ln n}{(1-p)^M} (1 + \delta_n)$ we have $\mathbb{E} X_{K'} < \epsilon_n$ for any positive integer n .

Note that (16) can be applied for sets T of size $t = K'$ whenever n is large enough since

$$K' = \frac{2 \ln n}{(1-p)^M} (1 + \delta_n) = \frac{2 \ln n}{e^{-mp^2 - O(px + mp^3)}} (1 + \delta_n) \leq K(1 + \delta'_n) < 2K$$

for some $\delta'_n \rightarrow 0$ since $px \rightarrow 0$ and $mp^3 \ll mp^2 p \ln n \ll Kp \rightarrow 0$.

So by (16) the expected number $Y_{K'}$ of K' -sets $T' \subseteq V$ such that $A_{T'}$ holds in G satisfies $\mathbb{E} Y_{K'} \leq \mathbb{E} X_{K'} < \epsilon'_n$ for some $\epsilon'_n \rightarrow 0$.

Let $B = B(n)$ be the event that each $v \in V$ satisfies $||S_v| - mp| < x$. By (10) we have that $\mathbb{P}(\bar{B}) \rightarrow 0$. Observe that B and $Rainbow(G, R, K')$ imply that $Y_{K'} > 0$. Now define $\epsilon''_n = \frac{\epsilon'_n}{1 - \mathbb{P}(\bar{B})} + \mathbb{P}(\bar{B})$ and note that $\epsilon''_n \rightarrow 0$ as $n \rightarrow \infty$. Since $Y_{K'}$ are non-negative by Markov's inequality we finally obtain:

$$\mathbb{P}(Rainbow(G, R, K')) \leq \mathbb{P}(Y_{K'} > 0 | B) + \mathbb{P}(\bar{B}) \leq \frac{\mathbb{E} Y_{K'}}{1 - \mathbb{P}(\bar{B})} + \mathbb{P}(\bar{B}) \leq \epsilon''_n.$$

It remains to show (16). We have

$$\mathbb{P}(A_T) = \prod_{i=1}^t \mathbb{P}(A_T(i) | A_T(1), \dots, A_T(i-1)).$$

Recall that $T = \{v_1, \dots, v_t\}$ and for $i = 1, \dots, t$ let s_i be the number of $j \in [i-1]$ such that $v_j v_i \notin R$. Fix $i \in [t]$. Assume without loss of generality, that $v_i v_1, \dots, v_i v_{s_i} \notin R$ and $v_i v_{s_i+1}, \dots, v_i v_{i-1} \in R$.

Since $tM = O(pKm)$

$$\sum_{j=1}^{s_i} |\bar{S}_{v_j}| \leq tM \ll m.$$

By conditioning on sizes of sets \bar{S}_{v_j} , $j = 1, \dots, s_i$ and the size of the set S_i by Lemma 2.10 we obtain that for large enough n

$$\mathbb{P}(A_T(i) | \bar{S}_{v_1} = A_1, \dots, \bar{S}_{v_{s_i}} = A_{s_i})$$

is maximised when the sets A_1, \dots, A_{s_i} are independent.

Therefore we can remove the conditioning so that

$$\mathbb{P}(A_T(i) | A_T(1), \dots, A_T(i-1)) \leq (1 - (1-p)^M)^{s_i}$$

and

$$\mathbb{P}(A_T) \leq (1 - (1-p)^M)^{\sum_i s_i} = (1 - (1-p)^M)^{\binom{t}{2} - |R_T|}.$$

□

Lemma 2.12 *Consider a sequence of random binomial intersection graphs $(G(n, m, p), n = 1, 2, \dots)$, $m = m(n)$, $p = p(n)$. Assume that the following holds:*

- 1) $np = O(1)$,
- 2) $\ln n \ll mp$,
- 3) $mp^2 \rightarrow \infty$,
- 4) $m(np)^3 \ll e^{2mp^2} (\ln n)^2$,

5) $e^{mp^2} p \ln n \rightarrow 0$.

Then there is a sequence $\delta_n \rightarrow 0$ such that

$$\mathbb{P}\left(\omega(G(n, m, p)) > 2e^{mp^2} \ln n(1 + \delta_n)\right) \rightarrow 0.$$

Proof Recall that T_w denotes the set of vertices that chose the key $w \in W$. We have $|T_w| \sim \text{Binom}(n, p)$ and

$$\mathbb{P}(|T_w| \geq k) \leq \binom{n}{k} p^k.$$

Call a key w red if $|T_w| \geq 3$. Denote by R the set of edges generated by all the red keys. Then since $\limsup np < \infty$

$$\mathbb{E} |R| \leq m \sum_{k=3}^{\infty} \binom{k}{2} \binom{n}{k} p^k \leq m(np)^2 (e^{np} - 1)/2 = O(m(np)^3).$$

Let $r = r(n)$ be such that $m(np)^3 \ll r \ll e^{2mp^2} (\ln n)^2$.

Consider n as fixed and fix a subset of attributes $W_0 \subseteq W$ and family $\mathcal{T}_0 = \{T_0(w) : w \in W_0\}$ of subsets of V where each subset has size at least 3. Suppose

$$\sum_{w \in W_0} \binom{|T_0(w)|}{2} \leq r.$$

Consider the event $A(W_0, \mathcal{T}_0)$ that in $G = G(n, m, p)$ the set of red attributes is exactly W_0 and $T_w = T_0(w)$ for each $w \in W_0$.

The sets T_w , $w \in W$ are iid with a common distribution T so that $|T| \sim \text{Binom}(n, p)$. Given that $|T| = k$ the set T is uniformly random from all the k -sets of V .

Conditional on the event $A(W_0, \mathcal{T}_0)$, the sets T_z for the remaining attributes $z \in W \setminus W_0$ are also iid and for $k \in \{0, 1, 2\}$

$$\mathbb{P}(|T_z| = k | A(W_0, \mathcal{T}_0)) = \frac{\mathbb{P}(|T_z| = k)}{\mathbb{P}(|T_z| \leq 2)}.$$

Furthermore, the size of the set $|T_z|$ given the event $A(W_0, \mathcal{T}_0)$ is stochastically dominated by $|T|$ (unconditional).

Let R_0 be the set of edges covered by the sets in \mathcal{T}_0 and write $K = 2e^{mp^2} \ln n(1 + \delta_n)$ where δ_n obtained in Lemma 2.11 for m, n, p and r .

Now an important observation is that

$$\mathbb{P}(\omega(G) \geq K | A(W_0, \mathcal{T}_0)) \leq \mathbb{P}(\text{Rainbow}(G, R_0, K)).$$

(Here the left side is equal to the probability that $G - R_0$ contains a rainbow copy of H such that $\omega(H \cup R_0) \geq K$, while the graph on the right side has (stochastically)

bigger sets $T_0(z)$ and the rainbow graph H can additionally have edges generated by attributes from W_0 .)

Therefore we have

$$\begin{aligned}
\mathbb{P}(\omega(G) \geq K) &= \sum_{W_0, \mathcal{T}_0} \mathbb{P}(\omega(G) \geq K | A(W_0, \mathcal{T}_0)) \mathbb{P}(A(W_0, \mathcal{T}_0)) \\
&\leq \sum_{W_0, \mathcal{T}_0: |R_0| \leq r} \mathbb{P}(\omega(G) \geq K | A(W_0, \mathcal{T}_0)) \mathbb{P}(A(W_0, \mathcal{T}_0)) + \mathbb{P}(|R| > r) \\
&\leq \sum_{W_0, \mathcal{T}_0: |R_0| \leq r} \mathbb{P}(\text{Rainbow}(G, R_0, K)) \mathbb{P}(A(W_0, \mathcal{T}_0)) + \mathbb{P}(|R| > r) \\
&\leq \max_{|R_0| \leq r} \mathbb{P}(\text{Rainbow}(G, R_0, K)) + \mathbb{P}(|R| > r).
\end{aligned}$$

When $n \rightarrow \infty$ the second term in the last line tends to zero by Markov's inequality. The first term tends to zero by Lemma 2.11 and the proof is complete. \square

Finally, we are ready to prove Lemma 2.3.

Proof of Lemma 2.3 Fix $\epsilon, \epsilon_c > 0$ such that $\epsilon < \epsilon_c < \min\{\epsilon_0, 1 - \alpha/2, (\beta - 2 + \alpha)/4 - 3\alpha\epsilon_1\}$ (recall that ϵ_0 is the constant from (2)). Let $\theta'_2 = ((1 - \alpha/2 - \epsilon_c)m \ln n)^{1/2}$. Let G'_1 be the subgraph of G_1 induced by vertices v with $X_v \leq \theta'_2$. First note that

$$D = |V(G_1) \setminus V(G'_1)| = c(\epsilon_c)K(1 + o_P(1)). \quad (17)$$

where $c(\epsilon_c) = (1 - \alpha/2 - \epsilon_c)^{-\alpha/2} - (1 - \alpha/2)^{-\alpha/2}$.

Indeed, $D \sim \text{Binom}(n, \mathbb{P}(X > \theta'_2) - \mathbb{P}(X > \theta_2))$, (2) and Lemma 2.5 yield that $\mathbb{E} D \sim c(\epsilon_c)K$ and (17) follows by a Chernoff bound, see (10). Since $c(\epsilon_c) \rightarrow 0$ as $\epsilon_c \rightarrow 0$ it suffices to show that for any $\epsilon_c > 0$

$$\omega(G'_1) = o_P(K).$$

We have that $|V(G'_1)|$ is stochastically dominated by $\text{Binom}(n, \mathbb{P}(X > \theta_1))$ and by (2) and a Chernoff bound whp $|V(G_1)| \leq n_1 = (1 + \epsilon)n^{1-\alpha/2+\alpha\epsilon_1}$. Let

$$p = (1 - \alpha/2 - \epsilon)^{1/2} \left(\frac{\ln n}{m} \right)^{1/2}$$

then by (10) in the random binomial intersection graph $G(n_1, m, p)$ all the sets S_v , $v \in [n_1]$ are of size at least θ'_2 whp. Therefore

$$\mathbb{P}(\omega(G'_1) \geq x) \leq \mathbb{P}(\omega(G(n_1, m, p)) \geq x) + o(1).$$

Now by Lemma 2.12 (we check that all of the conditions of this lemma hold) there is a sequence of positive numbers $\delta_n, \delta_n \rightarrow 0$ such that

$$\mathbb{P}(\omega(G(n_1, m, p)) > 2n^{1-\alpha/2-\epsilon} \ln n(1 + \delta_n)) \rightarrow 0 \quad (18)$$

which completes the proof. \square

3 Finite variance

In this section we prove Theorem 1.2.

3.1 Large cliques and rainbow K_4 s

Let $G = G(V, W)$ be an intersection graph of a family $\{S_v : v \in V\}$ of subsets of an attribute set W . Fix a graph H . We say that $S \subseteq V$ is a *witness of rainbow H* in G if there is a bijection f from $V(H)$ to S such that each edge $xy \in E(H)$ can be assigned a distinct element $w_{xy} \in W$ with $w_{xy} \in S_{f(x)} \cap S_{f(y)}$.

For a family \mathcal{C} of subsets of W , indexed by elements of V there is a unique dual family \mathcal{C}^* of subsets of V , indexed by elements of W given by $\mathcal{C}^* = \{S_v : v \in V\}$ where $S_v = \{w \in W : v \in C_w\}$. The family of sets $\{S_v : v \in V\}$ defining an intersection graph $G = G(V, W)$ is dual to the family of its monochromatic cliques $\{T_w : w \in W\}$.

Given a family \mathcal{C} of subsets of V indexed by W and a graph H , we call a set $S \subseteq V$ a *witness of rainbow H* , if S is a witness of rainbow H in the intersection graph defined by the dual family \mathcal{C}^* (for example, if H is a complete graph, this means that each pair of elements in S can be assigned a different set in \mathcal{C} that covers that pair). We call \mathcal{C} a *clique cover* of H , if $V(H) \subseteq V$ and for each edge $xy \in E(H)$ there is a set in \mathcal{C} containing both x and y .

The main lemma of this section is the following structural result for clique covers.

Lemma 3.1 *Let K and p be positive integers. There are positive integers $h = h(K)$ and $t_0 = t_0(K, p)$ such that the following holds.*

Let W be a set of size at least $t + h$ for some positive integer $t \geq t_0 - h$. Every clique cover $\mathcal{C} = \{C_1, \dots, C_r\}$ of the complete graph on W with $\max_i |C_i| \leq t$ and $\max_{i \neq j} |C_i \cap C_j| \leq p$ contains at least K witnesses of rainbow K_4 .

Furthermore, we may take $h(1) = 4$, $h(k) = O(k^{1/4})$ and $t_0(k, p) = O(p(k^{1/2} + p))$.

Proof Let h be the smallest integer such that $\binom{h}{4} \geq K$. Suppose that $|W| \geq t + h$ and \mathcal{C} has fewer than K witnesses of rainbow K_4 . Note that \mathcal{C} must not contain a rainbow K_h , for otherwise there are at least $\binom{h}{4} \geq K$ witnesses of rainbow K_4 . By Lemma 2.8 if \mathcal{C} does not contain a rainbow K_h then there is a constant $a = a(K)$ ($a = \frac{\ln h}{ch^3}$ where c is the constant from Lemma 2.8) such that $|B| \geq a|W|$ for some set $B \in \mathcal{C}$. Now $|B| \leq t$ so $W \setminus B$ contains a set S of size h , say $S = \{x_1, \dots, x_h\}$. If every pair of vertices $\{x_i, x_j\}$ in the set S , $1 \leq i < j \leq h$ is covered by at least 6 sets of \mathcal{C} then for any 4-subset S' of S we can assign a unique set for each pair of elements in S' and demonstrate a rainbow K_4 . In that case \mathcal{C} contains at least $\binom{h}{4} \geq K$ witnesses of rainbow K_4 .

So suppose one of the pairs of elements from S is contained in at most 5 sets in \mathcal{C} and without loss of generality assume it is $\{x_1, x_2\}$. Let i_{x_1, x_2} be the index of one of those sets (at least one set must cover both x_1 and x_2 as \mathcal{C} is a clique cover of the complete graph).

Since the intersection of any two distinct sets in \mathcal{C} has at most p elements, in order for every pair $\{\{x_1, y\}, y \in B\}$ to be covered at least $\frac{|B|}{p}$ sets need to be used. Thus we can pick a set $B_1 \subseteq B$ of size at least $\frac{|B|}{p}$ where each $y \in B_1$ is represented by a unique set $C_{i_{x_1, y}}$ containing both x_1 and y .

Now remove from B_1 the elements y such that $x_2 \in C_{i_{x_1, y}}$ (there are at most 5 of them). Also remove those elements y which belong to the set $C_{i_{x_1, x_2}}$ (there are at most p of them). Call the newly formed set B'_1 . Notice that

$$|B'_1| \geq \frac{|B|}{p} - 5 - p \geq \frac{a|W|}{p} - 5 - p.$$

Let us now treat the indices $\{1, \dots, r\}$ of the sets in \mathcal{C} as colours. Consider the complete graph G on vertices $B \cup \{x_1, x_2\}$. Colour each edge $x_1 y$ of G , where $y \in B'_1$ with the colour $i_{x_1, y}$. Colour the edge $x_1 x_2$ with i_{x_1, x_2} and for every edge $y_1 y_2 \in B'_1$ use the colour $i_{y_1, y_2} = j$ corresponding to the set B (so that B is the set C_j).

Finally let $\{x_2, y_2\}$ be coloured with an arbitrary colour i_{x_2, y_2} such that $C_{i_{x_1, y_2}}$ covers both x_2 and y_2 .

Take any element $y_1 \in B'_1$. We claim that for any element $y_2 \in B'_1 \setminus C_{i_{x_2, y_2}}$, the set $\{x_1, x_2, y_1, y_2\}$ witnesses rainbow K_4 . Indeed, by the construction, the colour i_{x_1, x_2} of the edge $x_1 x_2$ occurs only once, because $B'_1 \cap C_{i_{x_1, x_2}} = \emptyset$. Similarly, $x_1, x_2 \notin B$, so j occurs only once. The colours of the two other edges incident to x_1 occur only once, since we removed all candidates y such that $x_2 \in C_{i_{x_1, y}}$. Finally, $i_{x_2, y_1} \neq i_{x_2, y_2}$ since we chose y_2 outside $C_{i_{x_2, y_1}}$.

How many such witnesses can we form? For any y_1 we could choose $|B'_1| - |B'_1 \cap C_{i_{x_1, y_1}}| \geq |B'_1| - p$ suitable y_2 . Repeating this for every y_1 we will produce every 4-set at most twice. So \mathcal{C} contains at least

$$\frac{|B'_1|(|B'_1| - p)}{2} \geq \frac{1}{2} \left(\frac{a|W|}{p} - 5 - 2p \right)^2$$

witnesses of rainbow K_4 .

Thus since \mathcal{C} has fewer than K witnesses of rainbow K_4 and $|W| \geq t + h$ it must be that

$$\frac{1}{2} \left(\frac{a|W|}{p} - 5 - 2p \right)^2 < K$$

or

$$|W| < \frac{p}{a} \left(\sqrt{2K} + 5 + 2p \right).$$

Therefore we can set

$$t_0 = \frac{p}{a} \left(\sqrt{2K} + 5 + 2p \right)$$

to complete the proof. \square

The random power-law graph studied by Janson, Luczak and Norros [9] whp does not contain K_4 as a subgraph if the exponent α is larger than 2. In our case a similar result holds for rainbow cliques.

Lemma 3.2 *Let $(G(n), n = 1, 2, \dots)$ be a sequence of random intersection graphs such that $\mathbb{E}Y(n)^2 = O(1)$. Then the number $R = R(n)$ of 4-sets $S \subseteq V(G(n))$ that witness a rainbow K_4 in $G(n)$ satisfies*

$$\mathbb{E} R \leq \frac{(\mathbb{E} Y^2)^4}{4!} = O(1).$$

Furthermore, if there is a sequence $\epsilon_n \rightarrow 0$ such that $n\mathbb{P}(Y(n) \geq \epsilon_n n^{1/2}) \rightarrow 0$ then $G(n)$ does not contain a rainbow K_4 whp.

The relevant probability for a fixed graph is given by

Lemma 3.3 *Let $G = G(k, m, P)$ be a random intersection graph. Let X_1, \dots, X_k be the sizes of random subsets of G . For any integers x_1, \dots, x_k such that $\mathbb{P}(X_1 = x_1, \dots, X_k = x_k) > 0$*

$$\mathbb{P}(G \text{ has a rainbow } K_k | X_1 = x_1, \dots, X_k = x_k) \leq m^{-\frac{k(k-1)}{2}} (x_1 x_2 \dots x_k)^{k-1}.$$

Proof By considering all injective maps assigning for each edge $xy \in E(K_k)$ an attribute $w_{xy} \in S_x \cap S_y$ and using the union bound and independence, we obtain that the relevant conditional probability is at most

$$\binom{m}{k} \frac{\binom{x_1}{k-1}}{\binom{m}{k-1}} \frac{\binom{x_2}{k-1}}{\binom{m}{k-1}} \dots \frac{\binom{x_k}{k-1}}{\binom{m}{k-1}} \leq \frac{(x_1 x_2 \dots x_k)^{k-1}}{m^{k(k-1)/2}}.$$

□

Proof of Lemma 3.2 in $G(n)$. Denote $X_v = X_v(n) = |S_v(n)|$ and $Y = Y(n)$, also write $x \wedge y = \min(x, y)$. We have, using Lemma 3.3

$$\mathbb{E} R \leq \binom{n}{4} \mathbb{E} \left(\frac{X_1 X_2 X_3 X_4}{m^6} \wedge 1 \right) \leq \frac{n^4}{4!} \frac{\mathbb{E} (X_1 X_2 X_3 X_4)^2}{m^4} = \frac{(\mathbb{E} Y^2)^4}{4!}.$$

For the second part of the lemma, let $b = b(n) = \epsilon_n \sqrt{m}$ and let $A = A(n)$ be the event that $\max_{v \in V} X_v \leq b$. Then

$$\mathbb{P}(R \geq 1) \leq \mathbb{E} R \mathbf{I}_A + \mathbb{P}(A). \quad (19)$$

By the union bound the second term is at most

$$n\mathbb{P}(X > b) = n\mathbb{P}(Y > \epsilon_n n^{1/2}) \rightarrow 0.$$

The first term by Lemma 3.3 satisfies

$$\begin{aligned} \mathbb{E} R \mathbf{I}_A &\leq \binom{n}{4} m^{-6} \mathbb{E} (X_1 X_2 X_3 X_4)^3 \mathbf{I}_A \leq \frac{(\mathbb{E} X^2)^4 n^4 b^4}{4! m^6} \\ &= \epsilon_n^2 (\mathbb{E} Y^2)^4 \rightarrow 0. \end{aligned}$$

So $\mathbb{P}(R \geq 1) \rightarrow 0$. □

Given a random intersection graph G , we let $W(G)$ denote its set of attributes. The next result shows that the structure of random intersection graphs with $\mathbb{E}Y(n)^2 = O(1)$ is relatively simple.

Lemma 3.4 *Let $(G(n), n = 1, 2, \dots)$, $G(n) = G(n, m(n), P(n))$ be a sequence of random intersection graphs such that $\mathbb{E}Y(n)^2 = O(1)$. If $m(n) \rightarrow \infty$ then whp each pair $\{w', w''\} \subseteq W(G(n))$ belongs to at most two random sets S_v , $v \in V(G(n))$.*

Equivalently, this lemma says that the intersection of any two monochromatic cliques of $G(n)$ is of size at most 2 whp.

Proof For $G = G(n)$ fix a pair of distinct elements $w', w'' \in W(G)$ and $v \in V(G)$. We have

$$\begin{aligned} \mathbb{P}(w', w'' \in S_v) &= \sum_{k=0}^m \mathbb{P}(|S_v| = k) \frac{k(k-1)}{m(m-1)} = \frac{\mathbb{E}X^2 - \mathbb{E}X}{m(m-1)} \\ &\leq \frac{\mathbb{E}Y^2}{n(m-1)} \leq \frac{c}{nm}. \end{aligned}$$

for some constant $c > 0$.

By the union bound the probability that there is a pair of elements that is contained in at least k sets is at most

$$\binom{m}{2} \binom{n}{k} \mathbb{P}(w', w'' \in S_v)^k \leq m^2 \left(\frac{en}{k}\right)^k \left(\frac{c}{nm}\right)^k \leq m^2 \left(\frac{ec}{km}\right)^k,$$

which tends to zero when $k \geq 3$. \square

Proof of Theorem 1.3 Let $R = R(n)$ be the number of 4-sets $S \subseteq V(G(n))$ witnessing rainbow K_4 in $G(n)$. By Lemma 3.4, the intersection of any two monochromatic cliques has at most 2 vertices whp. In that case, by Lemma 3.1 either $\omega(G(n)) < t_0(R+1, 2)$ or $\omega(G) < \omega'(G) + h(R+1)$.

The proof of Lemma 3.1 shows that the functions $t_0(k, 2)$ and $h(k)$ are $O(\sqrt{k})$ as $k \rightarrow \infty$. Thus,

$$\omega(G(n)) \leq \omega'(G(n)) + Z(n)$$

where $Z(n) = t_0(R+1, 2) + h(R+1)$, and $\mathbb{E}Z_n^2 = O(\mathbb{E}R) = O(1)$ by Lemma 3.2.

If $n\mathbb{P}(Y(n) > \epsilon_n n^{1/2}) \rightarrow 0$ for some $\epsilon_n \rightarrow 0$ then by Lemma 3.2 $G(n)$ whp does not contain a rainbow K_4 , so whp $\omega(G) \leq t_0(1, 2) \vee (\omega'(G) + 3)$. \square

3.2 Monochromatic cliques and balls and bins

In this section we find the asymptotics of $\omega'(G(n, m, P))$. We start with a simple lemma for approximating the asymptotic behaviour of the maximum bin load $M(N, m)$.

Lemma 3.5 *Let $(N(n))$ and $(m(n))$ be sequences of positive integers such that $N = N(n) \rightarrow \infty$ and $m = m(n) \rightarrow \infty$. Let $(\delta_n), (\epsilon_n)$ be positive, $\delta_n \rightarrow 0, \epsilon_n = o(\delta_n)$. Write $M' = M'(n) = M(\lfloor N(1 + \epsilon_n) \rfloor, m)$ and $M = M(n) = M(N, m)$. Then there is a coupling between M' and M such that $M \leq M'$ and*

$$\mathbb{P}(M' - \delta_n \mathbb{E}M' \leq M) \rightarrow 1.$$

Proof We produce the coupling as follows. Throw $\lfloor N(1 + \epsilon_n) \rfloor$ balls into m bins. This gives an instance of M' . Denote by L the label of the bin with the lowest index realising the maximum.

Now delete a uniformly random set of $\lfloor \epsilon_n N \rfloor$ balls. The configuration with the remaining N balls gives an instance of $M \leq M'$. If $M' = M'_0$ and $L = L_0$, for some fixed M'_0, L_0 , the expected number of balls deleted from the bin L is by the union bound at most

$$\epsilon_n N \frac{M'_0}{N} = \epsilon_n M'_0.$$

Now applying Markov's inequality and averaging

$$\mathbb{P}(M' - M \geq t) \leq \epsilon_n t^{-1} \mathbb{E} M'$$

and the result follows by putting $t = \delta_n \mathbb{E} M'$. \square

Proof of Theorem 1.3 Write $G = G(n)$. The choice of random sets of sizes X_1, \dots, X_n in G can be interpreted in the balls and bins setting as follows. We have m distinct bins and for each i we have X_i balls labelled i . For each i we choose a random subset of X_i bins and place one ball labelled i to each of the X_i bins. Then T_w is the set of labels of the balls in the bin w .

We now relax the requirement that a bin can contain only one ball labelled i . Namely, for each ball let us select one bin independently at random. Call a ball *bad*, if it landed in the same bin with another ball with the same label. By conditioning and the union bound, for any label i

$$\mathbb{P}(\text{there is a bad ball labelled } i) = \mathbb{E} \binom{X_i}{2} \frac{1}{m^2} \leq \frac{\sup_n \mathbb{E} Y^2}{mn}.$$

Therefore, by the union bound

$$\mathbb{P}(\text{some ball is bad}) = O(m^{-1}). \quad (20)$$

Define $S'(i)$ as the set of bins that contain at least one ball labelled i . We have that $|S'(i)| \leq X_i$ are independent for $i \in [n]$ and given that $|S'(i)| = s$ each set of size s is equally likely (by symmetry). In particular, if $\tilde{M} = \tilde{M}(n)$ is the maximum number of balls contained in any single bin, for any n and any Borel set A we have

$$\mathbb{P}(\omega'(G) \in A) = \mathbb{P}(\tilde{M} \in A \mid \text{no bad balls});$$

and using (20)

$$d_{TV}(\omega'(G), \tilde{M}) = o(1).$$

Write $N = N(n) = \sum_{i=1}^n X_i$ and $M_0 = M_0(n) = M(\lfloor \mathbb{E} Y(mn)^{1/2} \rfloor, m)$. Since

$$|\mathbb{P}(\omega'(G) \in A) - \mathbb{P}(M_0 \in A)| \leq d_{TV}(\tilde{M}, M_0) + o(1),$$

to complete the proof it remains to prove the following.

$$d_{TV}(\tilde{M}, M_0) \rightarrow 0. \quad (21)$$

We will show this by constructing a pair of random variables \tilde{M}' and M'_0 , copies of \tilde{M} and M_0 respectively, in a single probability space so that

$$\mathbb{P}(\tilde{M}' \neq M'_0) \rightarrow 0.$$

Let (z_n) be a sequence of positive numbers, such that $z_n \rightarrow \infty$, $z_n = o((\ln m)^{1/2})$ and denote $\epsilon_n = z_n n^{-1/2}$. Define $N_- = \lfloor \mathbb{E} N(1 - \epsilon_n) \rfloor$, $N_+ = \lfloor \mathbb{E} N(1 + \epsilon_n) \rfloor$ and $N_0 = \lfloor \mathbb{E} Y(mn)^{1/2} \rfloor$.

For each n , define the following sequence of random variables, independent of N and \tilde{M} . Similarly as in the proof of Lemma 3.5, throw N_+ balls into m bins. Now successively pick one ball uniformly at random and delete it, repeat this $N_+ - N_-$ times. Denote by M_k , $k = N_-, \dots, N_+$, the number of balls in a bin with maximum load after $N_+ - k$ steps. The sequence $(M_{N_-}, \dots, M_{N_+})$ is non-decreasing and M_k is distributed as $M(k, m)$ by symmetry.

Now define

$$\tilde{M}' = \begin{cases} M_N & \text{if } N \in [N_-, N_+] \\ \tilde{M} & \text{otherwise.} \end{cases}$$

Note that \tilde{M}' is distributed as $M(N, m)$ and

$$\mathbb{P}(\tilde{M}' \neq M_{N_0}) \leq \mathbb{P}(N \notin [N_-, N_+]) + \mathbb{P}(M_{N_-} \neq M_{N_+}).$$

By Chebyshev's inequality, since $\mathbb{E} Y^2 = O(1)$ and $\mathbb{E} Y = \Theta(1)$

$$\mathbb{P}(N \notin [N_-, N_+]) \leq \frac{\text{Var}(N)}{\epsilon_n^2 (\mathbb{E} N)^2} \rightarrow 0. \quad (22)$$

By Lemma 3.5, we have that $\mathbb{P}(M_{N_-} \neq M_{N_+}) \rightarrow 0$ as soon as $\epsilon_n \mathbb{E} M_{N_+} \rightarrow 0$. But this is true. Let B_1 be the number of balls in the first bin (at step $k = 0$). For $t = t(n) = \epsilon_n^{-1}/z_n$ we have

$$\begin{aligned} \mathbb{E} M_{N_+} &\leq t + m N_+ \mathbb{P}(B_1 \geq t) \\ &\leq t + m N_+ \binom{N_+}{t} m^{-t} \\ &\leq t + m N_+ \left(\frac{e N_+}{t m} \right)^t \\ &\leq t + \exp \left\{ O(\ln n + \ln m - n^{1/2} z_n^{-2} \ln m) \right\} \\ &= o(\epsilon_n^{-1}). \end{aligned}$$

□

4 Algorithms for finding the largest clique

In the analysis of algorithms in this section we assume that graphs are represented by the adjacency list data structure. We have already used a version of the

GREEDY-CLIQUE algorithm of [9] in Section 2.2 to give a lower bound for the maximum clique. The algorithm below sorts vertices by their degrees (instead of sorting by the random subset sizes).

```

GREEDY-CLIQUE(G):
  Let  $v^{(1)}, \dots, v^{(n)}$  be  $V(G)$  sorted by their degrees, descending
   $M \leftarrow \emptyset$ 
  for  $i = 1$  to  $n$ 
    if  $v^{(i)}$  is adjacent to each vertex in  $M$  then
       $M \leftarrow M \cup \{v^{(i)}\}$ 
  return  $M$ 

```

Proposition 4.1 *Consider a sequence of random intersection graphs $(G(n))$ as in Theorem 1.1. Suppose (5) holds for some $\epsilon > 0$. The algorithm GREEDY-CLIQUE finds a clique of size $\omega(G(n))(1 - o_P(1))$ in time $O(n^2)$.*

Proof (Sketch) The running time bound is obvious, we only need to check that the algorithm returns a clique of the correct size. Fix any positive constants a, b with $a < 1/4$, where ϵ is given in (2). Let θ_2, e_1 be as in Section 2.1. Set

$$\tau = \tau(n) = \mathbb{E}Y((1 - \alpha/2) \ln n + e_1)n^{1/2}(1 + n^{-a}).$$

and $\tilde{\theta} = \theta_2(1 + b)$. Let $Q = Q(n)$ be the set of vertices in $V(G(n))$ with degree larger than τ and let $R = R(n)$ be the set of vertices u in $V(G(n))$ with $X(u) > \tilde{\theta}$.

We claim that whp Q does not contain any vertex u with $X_u < \theta_2$ and whp $R \subseteq Q$. We skip the proof of this, since it consists of standard applications of the Chernoff bounds and is very similar to the proof of the second part of Lemma 1.4.

Finally, applying a Chernoff bound again, we can obtain that $|Q| \geq |R| = (1 + b)^{-\alpha}K(n)(1 + O_P(1))$ and by Lemma 2.1, $\omega(G(n)[Q]) = (1 + b)^{-\alpha}K(n)(1 + O_P(1))$. Now the claim follows by Theorem 1.2 since we can make b arbitrarily small. \square

For sequences of random intersection graphs with bounded degree variance (as in Theorem 1.2 and Theorem 1.3) we need a different but also extremely simple algorithm.

```

MONO-CLIQUE(G):
   $M \leftarrow \emptyset$ 
  for  $uv \in E(G)$ 
     $C \leftarrow \{u, v\} \cup (\Gamma(u) \cap \Gamma(v))$ 
    if  $|C| > |M|$  and  $C$  is a clique then
       $M \leftarrow C$ 
  return  $M$ 

```

A slightly more robust version of MONO-CLIQUE, which still gives the same result would be to construct a maximal clique starting from each edge (without backtracking). It is difficult to imagine a simpler algorithm for finding large cliques.

Theorem 4.2 *Consider a sequence of random intersection graphs $(G(n))$ as in Theorem 1.3. The algorithm MONO-CLIQUE finds a clique of size $\omega(G(n)) - O_P(1)$ in expected time $O(n)$.*

We first need the following lemma.

Lemma 4.3 *Let $(G(n))$ be as in Theorem 1.3. Then whp either $G(n)$ is empty or there is a monochromatic clique T with $|T| = \omega'(G(n))$ and two vertices $u, v \in T$ such that $u \neq v$ and $|S(u) \cap S(v)| = 1$.*

Proof By Theorem 1.3 (and the coupling provided by its proof) it suffices to consider the N balls into m bins model, where X_1 balls have label 1, X_2 balls have label 2, etc., and $N = X_1 + \dots + X_n$. The balls, even those with the same label, are distinguishable. We denote the label of a ball b by $\text{label}(b)$. As before, for $v \in [n]$ let $S'(v)$ be the set of (indices of) bins with a ball labeled v . For $j \in [m]$, let Z_j be the number of balls in the bin j . Let L be the index of the (leftmost) bin containing maximum number of balls.

Let $A = A(n)$ be the event that either the bin L contains two balls with labels l_1, l_2 respectively such that $S'(l_1) \cap S'(l_2) = \{L\}$ or $Z_L \leq 1$. Since by (20) no bin contains two balls with the same label whp, in order to prove the lemma, it suffices to show that $\mathbb{P}(A) \rightarrow 1$. To do this, we define three other events for each n

$$\begin{aligned} B &= B(n) := \{|N - \mathbb{E} N| \leq \epsilon \mathbb{E} N\} \\ C &= C(n) := \left\{ \sum_{i=1}^n X_i^2 / N \leq c m^{1/2} n^{-1/4} \right\} \\ D &= D(n) := \left\{ \sum_{j=1}^m Z_j^2 \leq \frac{2\mathbb{E} N(\mathbb{E} N + m - 1)}{m} \right\}, \end{aligned}$$

where $c = 2\mathbb{E} Y^2 / \mathbb{E} Y$, and ϵ is any constant, $\epsilon \in (0; 0.1)$. We have

$$\mathbb{P}(\bar{A}) \leq \mathbb{P}(\bar{A}|B, C, D) + \mathbb{P}(\bar{B}) + \mathbb{P}(\bar{C}) + \mathbb{P}(\bar{D}).$$

Let us show first that the last three terms tend to 0. $\mathbb{P}(B) \rightarrow 0$ by (22) above.

Now consider the event C . Again, by (22)

$$\mathbb{P}(N < 2^{-1} m^{1/2} n^{1/2} \mathbb{E} Y) = \mathbb{P}(N < 0.5 \mathbb{E} N) \rightarrow 0.$$

Also, by Markov's inequality

$$\mathbb{P}\left(\sum_{i=1}^n X_i^2 > n^{1/4} m \mathbb{E} Y^2\right) \leq \frac{n \mathbb{E} X^2}{n^{1/4} m \mathbb{E} Y^2} = n^{-1/4}.$$

So combining the last two bounds we get

$$\mathbb{P}(\bar{C}) \leq \mathbb{P}(N < 2^{-1}m^{1/2}n^{1/2}\mathbb{E}Y) + \mathbb{P}\left(\sum_{i=1}^n X_i^2 > n^{1/4}m\mathbb{E}Y^2\right) \rightarrow 0.$$

To see that $\mathbb{P}(\bar{D}) \rightarrow 0$, note that since

$$\mathbb{P}(\bar{D}) \leq \mathbb{P}(\bar{D}|B) + \mathbb{P}(\bar{B}).$$

it suffices to show that

$$\mathbb{P}(\bar{D}|N = N_0) \rightarrow 0$$

for each sequence $N_0 = N_0(n)$ with $|N_0 - \mathbb{E}N| \leq \epsilon \mathbb{E}N$. A simple calculation yields that

$$\begin{aligned} \mathbb{E}\left(\sum_{j=1}^m Z_j^2\right) &= \frac{N_0(N_0 + m - 1)}{m} \\ \text{Var}\left(\sum_{j=1}^m Z_j^2\right) &= \frac{2(m-1)N_0(N_0 - 1)}{m^2} \end{aligned} \tag{23}$$

therefore by Chebyshev's inequality, since $N_0 \rightarrow \infty$, we have

$$\mathbb{P}(\bar{D}) \leq \frac{\text{Var}(\sum_{j=1}^m Z_j^2)}{((1-4\epsilon)\mathbb{E}\sum_{j=1}^m Z_j^2)^2} = o(m^{-1}).$$

It remains to prove that

$$\mathbb{P}(\bar{A}|B, C, D) \rightarrow 0. \tag{24}$$

For a fixed n and vectors $x \in \{0, 1, \dots, m\}^n$ and $z \in \{0, 1, \dots, mn\}^m$, such that $\sum_{i=1}^n x_i = \sum_{j=1}^m z_j$ we define events

$$E(x, z) := \{X_v = x_v, \text{ for each } v \in [n] \text{ and } Z_j = z_j \text{ for each } j \in [m]\}.$$

To prove (24), it is enough to show that for every pair (x, z) , such that $N_0 = \sum_{i=1}^n x_i = \sum_{j=1}^m z_j \in [(1-\epsilon)\mathbb{E}N; (1+\epsilon)\mathbb{E}N]$, $\sum_{i=1}^n x_i^2/N_0 \leq cm^{1/2}n^{-1/4}$, $\sum_{j=1}^m z_j^2 \leq 2m\mathbb{E}Z_1^2$, $\max_j z_j \geq 2$ we have, if n is large enough

$$\mathbb{P}(\bar{A}|E(x, z)) \leq 2m^{-1/2}n^{-1/2}(\mathbb{E}Y)^{-1} + 3c^2n^{-1/2}.$$

For each such x and z we may define a Markov chain M_{xy} with state space \mathcal{S}_{xy} consisting of all configurations of balls for which the event $E(x, z)$ occurs. Let the distribution of the initial state, s_0 , be specified by the random assignment of the balls conditional on the event $E(x, z)$. Then by symmetry s_0 is uniform over \mathcal{S}_{xy} .

The transition probabilities of M_{xy} are defined by a random walk in \mathcal{S}_{xy} as follows. Given that we are in state s choose two balls b_1, b_2 from the bin L independently without replacement, and choose another two balls b'_1, b'_2 from all N_0 balls

without replacement. Exchange the bins of b_1 and b'_1 , then exchange the bins of b_2 and b'_2 . This gives a new state s' of M_{xy} .

Note that if $b_2 \neq b'_1$ then the bin L in the state s' contains both balls b'_1 and b'_2 . Furthermore, by symmetry $\sum_s \mathbb{P}(s' \rightarrow s) = \sum_s \mathbb{P}(s \rightarrow s')$ for any state $s' \in \mathcal{S}_{xy}$. Therefore the uniform distribution over \mathcal{S}_{xy} is a stationary distribution for M_{xy} . So, denoting by \mathbb{P}_* the probability conditional on $E(x, z)$, we have

$$\mathbb{P}_*(\bar{A}) \leq \mathbb{P}_*(b'_1 = b_2) + \mathbb{P}_*(S'_{b'_1} \cap S'_{b'_2} \neq \emptyset); \quad \mathbb{P}_*(b'_1 = b_2) = \frac{1}{N_0};$$

and for all large enough n we have

$$\begin{aligned} \mathbb{P}_*(S'_{b'_1} \cap S'_{b'_2} \neq \emptyset) &= \sum_{i,j=1}^n \mathbb{P}_*(S'_i \cap S'_j \neq \emptyset) \mathbb{P}(\text{label}(b'_1) = i, \text{label}(b'_2) = j) \\ &\leq \sum_{i,j=1}^n \mathbb{P}_*(S'_i \cap S'_j \neq \emptyset) \frac{x_i x_j}{N_0^2} \\ &\leq \sum_{i,j=1}^n \left(x_i x_j \sum_{k=1}^m \frac{z_k^2}{N_0^2} \right) \frac{x_i x_j}{N_0^2} \end{aligned} \tag{25}$$

$$\leq \sum_{i,j=1}^n \frac{3x_i^2 x_j^2}{mN_0^2} = \frac{3(\sum x_i^2)^2}{mN_0^2} \tag{26}$$

$$\leq 3c^2 n^{-1/2}. \tag{27}$$

Here (25) follows since the conditional probability that a particular pair of balls lands in the same bin is by the union bound at most

$$\sum_{k=1}^m \frac{(z_k)_2}{(N_0)_2} \leq \frac{z_k^2}{N_0^2}.$$

The bound (26) follows since on the event $E(x, y)$, the events B and D occur, which imply, using (23) that for n large enough

$$\frac{\sum_k z_k^2}{N_0^2} \leq \frac{2}{(1-\epsilon)^2} \left(\frac{1}{m\mathbb{E}N} + \frac{1}{m^2} \right) \leq \frac{3}{m},$$

since $\mathbb{E}N \rightarrow \infty$ and $\epsilon < 0.1$.

Finally, the bound (27) follows since on $E(x, y)$ the event D occurs. \square

Proof of Theorem 4.2 The fact that MONO-CLIQUE finds a clique of order $\omega(G(n)) - O_P(1)$ (which is a maximum monochromatic clique) follows by Theorem 1.2 and Lemma 4.3.

The running time of the algorithm is the sum of the number of steps to find the common neighbours for each edge and the number of steps to test if these common neighbours form a clique.

We assume that the elements in each list in the adjacency list structure are sorted in increasing order. Otherwise they can be sorted using any standard sorting algorithm in $O(\sum_{v \in V} d(v)^2)$ time, where $V = V(G(n))$ and $d(v) = d_{G(n)}(v)$ is the degree of v in $G(n)$. The intersection of two lists of lengths k_1 and k_2 can be found in $O(k_1 + k_2)$ time, so the total time for finding common neighbours is

$$O\left(\sum_{v \in V} d(v)(d(v) - 1)\right).$$

The time for checking if pairs of vertices are connected is a constant times the number of 2-stars in $G(n)$. So the time for testing cliques is again $O(\sum_{v \in V} d(v)(d(v) - 1))$. But

$$\mathbb{E} \sum_{v \in V} d(v)(d(v) - 1) = O(n)$$

by (37) in the proof of Lemma 1.5. \square

5 Equivalence between set size and degree parameters

Proof of Lemma 1.4 We start by showing that if either $\mathbb{E}Y$ or $\mathbb{E}D_1$ converges and

$$\mathbb{E}Y \mathbf{I}_{Y > (an)^{1/2}} \rightarrow 0$$

for some positive $a = a(n) \rightarrow 0$ then

$$\mathbb{E}Y = (\mathbb{E}D_1)^{1/2} + o(1). \quad (28)$$

By the inclusion-exclusion principle we have

$$Z := 0 \vee \left(\frac{X_1 X_2}{m} - \frac{X_1^2 X_2^2}{m^2} \right) \leq \mathbb{P}(S_1 \cap S_2 \neq \emptyset | X_1, X_2) \leq \frac{X_1 X_2}{m}. \quad (29)$$

Notice that $\mathbb{E}Y = \Omega(1)$. This is clear if $\mathbb{E}Y \rightarrow y \in (0; \infty)$. Otherwise, $\mathbb{E}D_1 \rightarrow d$ and by (29)

$$\frac{(n-1)(\mathbb{E}Y)^2}{n} \geq \mathbb{E}D_1 = d + o(1).$$

Now observe that by the first inequality of (29)

$$\begin{aligned} \mathbb{E}Z &\geq \mathbb{E}Z \mathbf{I}_{X_1 X_2 \leq am} \geq (1-a)m^{-1} \mathbb{E}X_1 X_2 \mathbf{I}_{X_1 X_2 \leq am} \\ &\geq (1-a)m^{-1} \mathbb{E}X_1 \mathbb{E}X_2 - m^{-1} \mathbb{E}X_1 X_2 \mathbf{I}_{X_1 X_2 > am} \end{aligned} \quad (30)$$

and since $\mathbb{E}Y = \Theta(1)$ and $\mathbb{E}Y \mathbf{I}_{Y > (an)^{1/2}} \rightarrow 0$ we get

$$\begin{aligned} \mathbb{E}X_1 X_2 \mathbf{I}_{X_1 X_2 > am} &\leq \mathbb{E}X_1 X_2 (\mathbf{I}_{X_1 > (am)^{1/2}} + \mathbf{I}_{X_2 > (am)^{1/2}}) \\ &\leq 2\mathbb{E}X \mathbb{E}X \mathbf{I}_{X > (am)^{1/2}} \\ &= (m/n)^{1/2} \mathbb{E}X \mathbb{E}Y \mathbf{I}_{Y > (an)^{1/2}} = o((\mathbb{E}X)^2). \end{aligned}$$

Combining the last estimate with (30) and (29) we get that

$$\mathbb{P}(S_1 \cap S_2 \neq \emptyset) \sim \frac{(\mathbb{E} X)^2}{m},$$

and (28) follows. This completes the proof of the first claim.

Let $\tilde{\epsilon} \in (0; \epsilon)$ be smaller than ϵ_0 from (2) or ϵ' from (4), depending on which of the two conditions hold. To prove the second claim, we will show that for each $\delta \in (0; 1)$ and each sequence $(t_n, n = 1, 2, \dots)$ with $n^{1/2-\tilde{\epsilon}} \leq t_n \leq n^{1/2+\tilde{\epsilon}}$ we have

$$\mathbb{P}(Y_n \geq t_n) \geq (d^{1/2}(1+\delta))^{-\alpha} \mathbb{P}(D_1(n) \geq t_n) \quad (31)$$

and

$$\mathbb{P}(Y_n \geq t_n) \leq (d^{1/2}(1-\delta))^{-\alpha} \mathbb{P}(D_1(n) \geq t_n). \quad (32)$$

Proof of the lower bound (31). We prove by contradiction. Suppose there is a subsequence (n_k) and a sequence (b_k) with $n_k^{1/2-\tilde{\epsilon}} \leq b_k \leq n_k^{1/2+\tilde{\epsilon}}$ such that

$$\mathbb{P}(Y(n_k) \geq b_k) < (d^{1/2}(1+\delta))^{-\alpha} \mathbb{P}(D_1(n_k) \geq b_k) \quad \text{for } k = 1, 2, \dots$$

Let us define two sequences of positive numbers (t_k) and (l_k) depending on whether (2) or (4) holds. If (2) holds, then let $t_k = b_k$. If (4) holds, then let $l_k = b_k$. In each case, define the other sequence by $t_k = d^{1/2}(1+\delta/2)l_k$. Also write $Y_1(n) = (n/m)^{1/2}X_1(n)$. Then

$$\mathbb{P}(D_1(n_k) \geq t_k) = \mathbb{P}(D_1(n_k) \geq t_k, Y_1(n_k) \geq l_k) + \mathbb{P}(D_1(n_k) \geq t_k, Y_1(n_k) < l_k). \quad (33)$$

We claim that the first term on the right is bounded by

$$(c + o(1))\mathbb{P}(D_1(n_k) \geq t_k) \quad \text{as } k \rightarrow \infty,$$

where $c = \left(\frac{1+\delta/2}{1+\delta}\right)^\alpha < 1$. Indeed, if (4) holds then

$$\begin{aligned} \mathbb{P}(D_1(n_k) \geq t_k, Y_1(n_k) \geq l_k) &\leq \mathbb{P}(Y_1(n_k) \geq l_k) \\ &\leq (d^{1/2}(1+\delta))^{-\alpha} \mathbb{P}(D_1(n_k) \geq l_k) \sim c \mathbb{P}(D_1(n_k) \geq t_k). \end{aligned}$$

Similarly, if (2) holds then

$$\begin{aligned} \mathbb{P}(D_1(n_k) \geq t_k, Y_1(n_k) \geq l_k) &\leq \mathbb{P}(Y_1(n_k) \geq l_k) \\ &\sim d^{\alpha/2}(1+\delta/2)^\alpha \mathbb{P}(Y_1(n_k) \geq t_k) \leq c \mathbb{P}(D_1(n_k) \geq t_k). \end{aligned}$$

Let us now consider the second term on the right of (33). It satisfies

$$\mathbb{P}(D_1(n_k) \geq t_k, Y_1(n_k) < l_k) \leq \mathbb{P}(D_1(n_k) \geq t_k | X_1(n_k) = x_k), \quad (34)$$

where $x_k = \lfloor (m(n_k)/n_k)^{1/2} l_k \rfloor$. By (36), conditionally on $X_1(n_k) = x_k$, $D_1(n_k)$ is stochastically dominated by the random variable $B_k \sim \text{Binom}(n_k - 1, p_k)$ where

$p_k = \frac{x_k \mathbb{E} X}{m}$. Since $t_k = d^{1/2}(1 + \delta/2)l_k \sim (1 + \delta/2)\mathbb{E} B_k$, by the Chernoff bound (9) we have for large k

$$\mathbb{P}(D_1(n_k) \geq t_k | X_1(n_k) = x_k) \leq \mathbb{P}(B_k \geq (1 + \delta/3)\mathbb{E} B_k) \leq e^{-\Omega(\mathbb{E} B_k)} = e^{-\Omega(n_k^{1/2-\varepsilon})}.$$

Using the last estimate in (34), we finally have that for large k the right hand side of (33) is less than the left hand side, which is a contradiction. This finishes the proof of (31).

Proof of the upper bound (32). Suppose, that there is a subsequence (n_k) and a sequence (b_k) with $n_k^{1/2-\bar{\varepsilon}} \leq b_k \leq n_k^{1/2+\bar{\varepsilon}}$ such that

$$\mathbb{P}(Y(n_k) \geq b_k) > (d^{1/2}(1 - \delta))^{-\alpha} \mathbb{P}(D_1(n_k) \geq b_k), \quad k = 1, 2, \dots$$

Again, define sequences of positive numbers $(t_k), (l_k)$, depending on what we want to prove. If (2) holds then let $t_k = b_k$. If (4) holds, let $l_k = b_k$. Define the other sequence by setting $t_k = d^{1/2}(1 - \delta/2)l_k$. We have that

$$\mathbb{P}(D_1(n_k) \geq t_k) \geq \mathbb{P}(Y(n_k) \geq l_k) \mathbb{P}(D_1(n_k) \geq t_k | Y(n_k) \geq l_k). \quad (35)$$

By the definition of (l_k) and (t_k) and either (2) or (4), the first term on the right is at least $(c + o(1))\mathbb{P}(D_1(n_k) \geq t_k)$ where $c = \left(\frac{1-\delta/2}{1-\delta}\right)^\alpha > 1$. Therefore it suffices to show that the second term of (35) is $1 - o(1)$ to obtain a contradiction.

Let $x_k = \lfloor (m(n_k)/n_k)^{1/2} l_k \rfloor$. Conditionally on $X_1 = x_k$, $D_1(n_k)$ has distribution $\text{Binom}(n_k - 1, p_k)$, where p_k is the probability that a random subset of size x_k intersects a random subset of size X_2 .

To estimate p_k , notice that (29) and (30) still hold when X_1 and X_2 have different distributions ($\mathbb{P}(X_1(n_k) = x_k) = 1$ and $X_2(n_k)$ has distribution $P(n_k)$). Also, with $a(n) = (\ln n)^{-1}$ we have

$$\mathbb{E} x_k X_2(n_k) \mathbf{I}_{x_k X_2 > a(n_k) m(n_k)} \leq x_k \mathbb{E} X(n_k) \mathbf{I}_{Y(n_k) > a(n_k) n_k / l_k} = o(x_k \mathbb{E} X(n_k))$$

by (5). Therefore by (29) we get that

$$p_k \sim \frac{\mathbb{E} X(n_k) x_k}{m(n_k)} \sim \frac{\mathbb{E} Y(n_k) l_k}{n_k} \quad (36)$$

as $k \rightarrow \infty$.

So $t_k = d^{1/2}(1 - \delta/2)l_k \sim (1 - \delta/2)\mathbb{E} B_k$, where B_k is a random variable with distribution $\text{Binom}(n_k - 1, p_k)$. Using (9) again, we get for k large enough

$$\mathbb{P}(D_1(n_k) < t_k | Y(n_k) \geq l_k) \leq \mathbb{P}(B_k < (1 - \delta/3)\mathbb{E} B_k) = e^{-\Omega(\mathbb{E} B_k)} \rightarrow 0.$$

□

We now prove Lemma 1.5. Similar identities have been obtained in [4], we include the proof for completeness.

Proof of Lemma 1.5 The identity (7) follows from (28) since

$$\mathbb{E} Y \mathbf{I}_{Y > \epsilon_n n^{1/2}} \leq (\mathbb{E} Y^2 \mathbf{I}_{Y > \epsilon_n n^{1/2}})^{1/2} \rightarrow 0.$$

It remains to show (8). Note that $\mathbb{E} D_1(D_1 - 1) = 2\mathbb{E} V_1$, where V_1 is the number of 2-stars in $G(n)$ centered at vertex 1. If $A = A(n)$ is the event that $12, 13 \in E(G(n))$ then $\mathbb{E} V_1 = \binom{n}{2} \mathbb{P}(A)$. Conditional on sizes X_1, X_2, X_3 of the first three random subsets of $G(n)$, the events $12 \in E(G(n))$ and $13 \in E(G(n))$ are independent. Furthermore, by symmetry, $\mathbb{P}(A|X_1, X_2, X_3) = \mathbb{P}(A|S_1 = [X_1], X_2, X_3)$. By the inclusion-exclusion principle, for $j = 2, 3$

$$\frac{X_1 X_j}{m} \left(1 - \frac{X_1 X_j}{m} \right) \leq \mathbb{P}(1j \in E(G(n)) | X_1, X_2, X_3) \leq \frac{X_1 X_j}{m}.$$

So

$$\mathbb{P}(A) \leq \frac{\mathbb{E} X_1^2 X_2 X_3}{m^2} = \frac{\mathbb{E} Y^2 (\mathbb{E} Y)^2}{n^2}. \quad (37)$$

Since $\mathbb{P}(A|X_1, X_2, X_3)$ is nonnegative

$$\begin{aligned} \mathbb{P}(A) &\geq \mathbb{E} \left(\mathbb{P}(A|X_1, X_2, X_3) \mathbf{I}_{X_1 \leq \epsilon m^{1/2}} \mathbf{I}_{X_2 \leq \epsilon m^{1/2}} \mathbf{I}_{X_3 \leq \epsilon m^{1/2}} \right) \\ &\geq (1 - \epsilon^2)^2 \mathbb{E} \left(\frac{X_1^2 X_2 X_3}{m^2} \mathbf{I}_{X_1 \leq \epsilon m^{1/2}} \mathbf{I}_{X_2 \leq \epsilon m^{1/2}} \mathbf{I}_{X_3 \leq \epsilon m^{1/2}} \right) \\ &\geq (1 - \epsilon^2)^2 \mathbb{E} \left(\frac{X_1^2 X_2 X_3}{m^2} (1 - \mathbf{I}_{X_1 > \epsilon m^{1/2}} - \mathbf{I}_{X_2 > \epsilon m^{1/2}} - \mathbf{I}_{X_3 > \epsilon m^{1/2}}) \right) \\ &= \frac{\mathbb{E} Y^2 (\mathbb{E} Y)^2}{n^2} (1 - o(1)). \end{aligned}$$

The bound in the last line follows since $\mathbb{E} Y^2 \geq (\mathbb{E} Y)^2 = \Omega(1)$ and by the uniform convergence assumption

$$\mathbb{E} X_1^2 \mathbf{I}_{X_1 > \epsilon m^{1/2}} = \frac{m}{n} \mathbb{E} Y^2 \mathbf{I}_{Y > \epsilon n^{1/2}} = o(\mathbb{E} X^2)$$

and also for $j = 2, 3$

$$\mathbb{E} X_j \mathbf{I}_{X_j > \epsilon m^{1/2}} = m^{1/2} n^{-1/2} \mathbb{E} Y \mathbf{I}_{Y > \epsilon n^{1/2}} = o(\mathbb{E} X).$$

Now we have that $\mathbb{E} D_1(D_1 - 1) = \mathbb{E} Y^2 (\mathbb{E} Y)^2 (1 - o(1))$ and (8) follows by (7). \square

6 Concluding remarks

In this work we determined the order of the clique number in $G(n, m, P)$ for a wide range of $m = m(n)$ and $P = P(n)$. We saw that in sparse power-law random intersection graphs with unbounded degree variance, the clustering property of $G(n, m, P)$ has little influence in the formation of the maximum clique. This suggests that simpler models, such as the one in [9], may be preferable in the case of very heavy degree tails. However, when the degree variance is bounded, most random graph models, including the Erdős-Rényi graph and the model of [9] have only bounded size cliques whp. In contrast, we showed that in random intersection graphs the clique number can still diverge slowly.

We have a kind of “phase transition” as the tail index α for the random subset size (degree) varies, see (2). Assume, for example that $m = \Theta(n)$. When $\alpha < 2$, the random graph $G(n, m, P)$ whp contains cliques of only logarithmic size. When $\alpha > 2$, it whp contains a ‘giant’ clique of polynomial size. But what happens when (2) is satisfied with $\alpha = 2$ but the degree variance is unbounded?

We proposed a surprisingly simple algorithm for finding (almost) the largest clique in sparse random intersection graphs with finite degree variance. The performance of both GREEDY-CLIQUE and MONO-CLIQUE algorithms can be of further interest, since these algorithms do not use the possibly hidden random subset structure. How well would they perform on arbitrary sparse empirical networks? Can we suspect a hidden intersecting sets structure for networks where the MONO-CLIQUE algorithm performs well?

Another direction of possible future research would be to determine the asymptotic clique number in dense random intersection graphs. For example, even in the random uniform hypergraph case where $m = \Theta(n)$ and the random subset size $X(n) = \Omega(n^{1/2})$ is deterministic, exact asymptotics of the clique number remain open.

Acknowledgement. V. Kurauskas would like to thank prof. A. Iksanov for clarification of a few technical details.

References

- [1] N. Alon, T. Jiang, Z. Miller and D. Pritkin, Properly colored subgraphs and rainbow subgraphs in edge-colorings with local constraints, *Random Struct. Algorithms* **23** (2003), 409–433.
- [2] G. Bianconi and M. Marsili, Emergence of large cliques in random scale-free networks. *Europhys. Lett.* **74** (2006), 740–746.
- [3] M. Bloznelis, J. Jaworski and V. Kurauskas, Assortativity and clustering coefficient of sparse random intersection graphs, *submitted*, arXiv:1209.4675.
- [4] M. Bloznelis, Degree and clustering coefficient in sparse random intersection graphs, *Ann. Appl. Probab.*, to appear.
- [5] M. Bloznelis, Degree distribution of a typical vertex in a general random intersection graph, *Lithuanian Math. J.* **48** (2008) 38–45.
- [6] M. Deijfen and W. Kets, Random intersection graphs with tunable degree distribution and clustering, *Probab. Eng. Inf. Sci.* **23** (2009) 661–674.
- [7] J. Galambos and E. Seneta, Regularly varying sequences, *Proc. Amer. Math. Soc.* **41**(1973) 110–116.
- [8] E. Godehardt and J. Jaworski, Two models of random intersection graphs for classification. In: O. Optiz and M. Schwaiger, Editors, *Studies in Classification, Data Analysis and Knowledge Organization* **22**, Springer, Berlin (2003), 67–82.

- [9] S. Janson, T. Łuczak and I. Norros, Large cliques in a power-law random graph, *J. Appl. Probab.* **47** (2010), 1124–1135.
- [10] S. Janson, T. Łuczak, and A. Ruciński, Random Graphs, *Wiley-Interscience Series in Discrete Mathematics and Optimization*, Wiley-Interscience, New York, 2000.
- [11] M. Karoński, E.R. Scheinerman and K.B. Singer-Cohen, On random intersection graphs: the subgraph problem, *Comb. Probab. Comput.* **8** (1999) 131–159.
- [12] V. Kolchin, B. Sevstyanov and V. Chistyakov, Random Allocations, *V.H. Winston and Sons*, 1978, Washington D.C.
- [13] C. McDiarmid, Concentration, in *Probabilistic Methods for Algorithmic Discrete Mathematics*, M. Habib, C. McDiarmid, J. Ramirez-Alfonsin and B. Reed Eds., Springer, New York (1998) 195–248.
- [14] K. Rybarczyk and D. Stark, Poisson approximation of the number of cliques in random intersection graphs, *J. Appl. Probab.*, **47** (2010), 826–840.